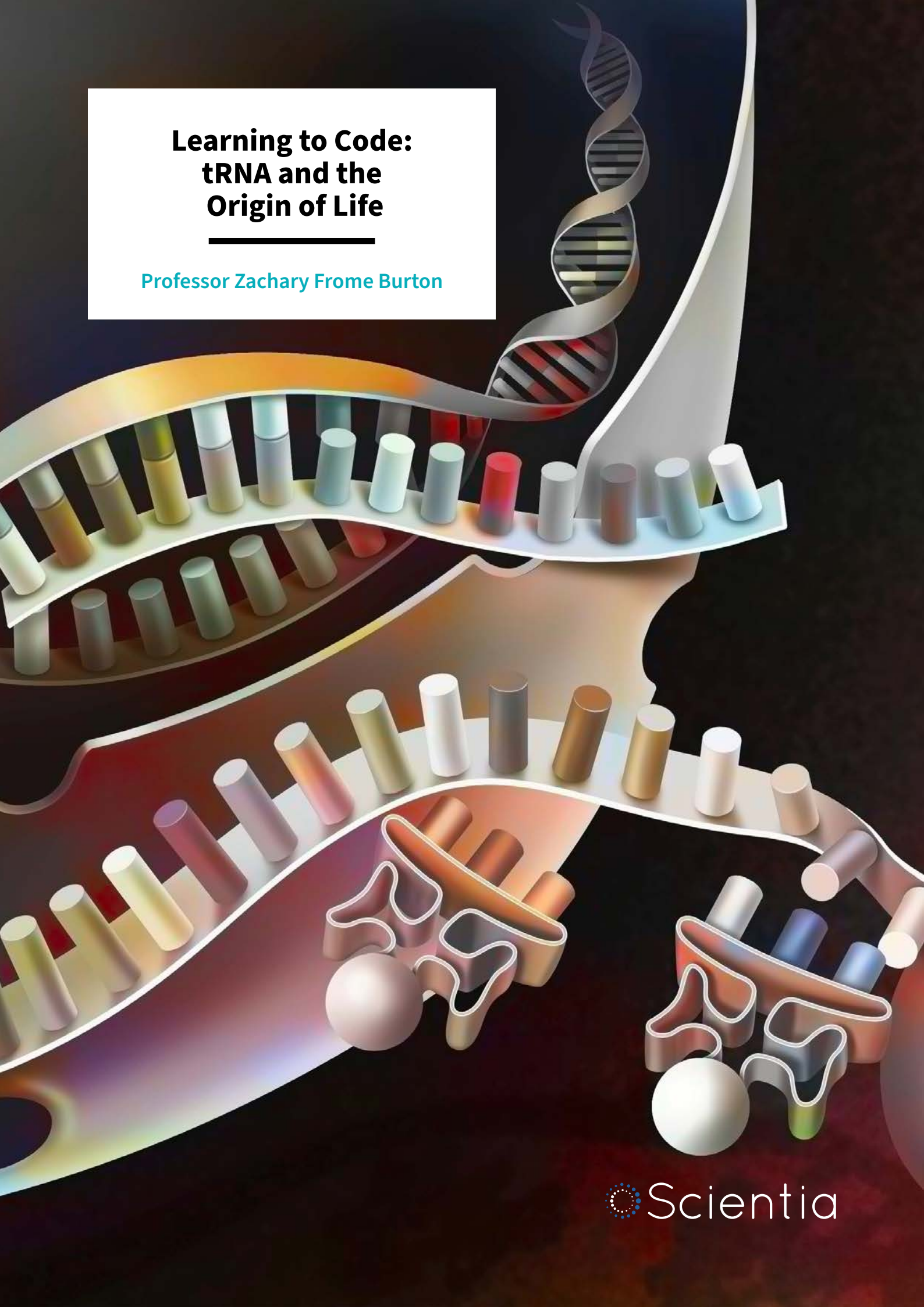# Learning to Code: tRNA and the Origin of Life

Professor Zachary Frome Burton

LEARNING TO CODE: tRNA AND THE ORIGIN OF LIFE

The importance of tRNA goes beyond its role in protein translation. tRNA is the molecule that 'learned to code' – its primordial emergence led to the genetic code and life itself. **Professor Zachary Frome Burton** of Michigan State University and Professor Lei Lei of the University of New England have reportedly solved tRNA evolution – proposing a theorem that sheds new light on the origin of life.

The Importance of tRNA is Not Lost in Translation

tRNA (transfer RNA) is well-known for its role as an adaptor in the translation of proteins. However, its significance extends far beyond this. tRNA originated in the pre-life era, 'learned to code', and played a pivotal role in the emergence of the genetic code. Professor Zachary Frome Burton (Michigan State University) and Professor Lei Lei (University of New England) have reportedly solved the problem of tRNA evolution by putting forward a simple yet elegant theorem with profound implications for evolutionary science.

tRNA's famous 'cloverleaf' secondary structure consisting of stems and loops, and its L-shaped tertiary structure, are the result of its unique evolutionary journey. The secondary structure consists of an acceptor stem formed by pairing between the 5' terminus and the 3' end, D loop, anticodon loop, V (variable) loop, and T loop. Interactions between the T loop and D loop form the elbow of the L-shaped molecule. Most tRNAs are type I with a shorter V chain (5 nt) compared to type II tRNAs (14 nt). Living organisms contain a diversity of tRNAs, collectively known as the tRNAome.

### The 4-Billion-Year-Old Message in a Bottle

The emergence of life about 4 billion years ago remains a fascinating mystery in evolutionary biology. Long before cells, proteins, or even DNA, there were tRNA precursors that formed simple self-replicating systems. This pre-life era has been divided into the polymer world, minihelix world, and tRNA world. Bottom-up strategies for investigating this history involve generating in vitro replicating chemical systems but may lead to evolutionary dead-ends.

Professors Burton and Lei take a top-down strategy. They carry out tRNA sequence analyses to explore preserved sequences originating in pre-life – like a 4-billion-year-old message in a bottle! They found that pre-life tRNA precursor sequences were highly ordered, consisting of RNA repeats and inverted repeats. 'This was a surprise,' Professor Burton reflects, 'as tRNA evolution was thought to be chaotic'. Conveniently for their analyses, this meant that pre-life sequences could be determined with near certainty.

The Mother of All tRNAs

Professors Burton and Lei have put forward a theorem explaining the origin of tRNA – the 3 31 nucleotide (nt) minihelix theorem. According to this, three 31-nt minihelices – the D, anticodon, and T loops – were joined together by ligation to form a 93 nt precursor molecule. The D loop minihelix was a 7 nt GCG repeat (5' acceptor stem – GCGGCGG) linked to a 17 nt UAGCC repeat (D loop minihelix core – UAGCCUAGCCUAGCCUA) linked to a 7 nt CGC repeat (3'-acceptor stem – CCGCCGC). The anticodon loop minihelix and T loop minihelix were probably originally identical – consisting of a 7 nt GCG repeat (5' acceptor stem – GCGGCGG) linked to a 17 nt stem-loop-stem (CCGGG_CU/???AA_CCCGG [/ indicates a U-turn and ? indicates an unknown base, either A, C, G, or U]) linked to a 7 nt CGC repeat (3'-acceptor stem – CCGCCGC). The 93 nt precursor underwent one or two deletions to form type II or type I tRNAs, respectively. Both type II and I tRNAs underwent a 9 nt deletion (removing 7 nt of the 3'-acceptor stem and 2 nt of the 5'-acceptor stem), while type I tRNA underwent an additional 9 nt deletion (removing 7 nt of the 5'-acceptor stem and 2 nt of the 3'-acceptor stem).

tRNA
(transfer RNA)
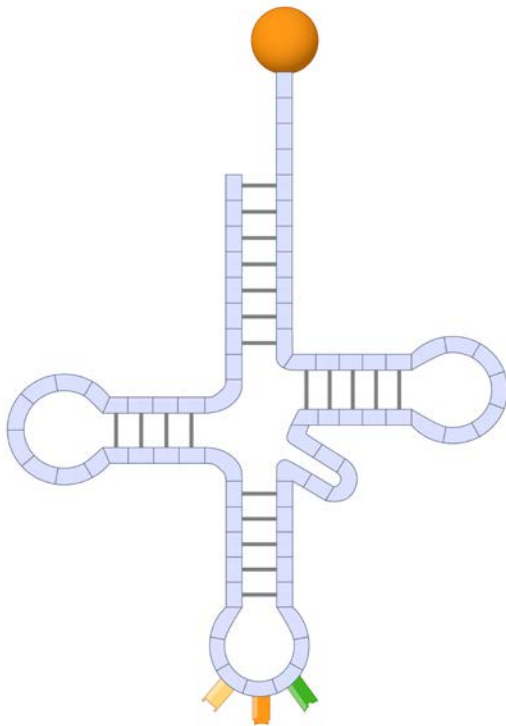
Helps to decode information
present in mRNA sequences
into proteins

Theorems are rare in biology – but Professors Burton and Lei are convinced that available data confirms theirs. Their theorem is based on a model of divergent evolution (divergence of tRNAs from a common ancestor) and is strongly supported by statistical analysis. Alternative models, based on convergent evolution or accretion (adding nucleotides) have been proposed, although these fall short. Convergent models assume that multiple small tRNA fragments converge upon the same homologous sequence. Similarly, accretion models assume that sequential addition of nucleotides will result in sequence homology. However, these are statistically less likely than divergent tRNA evolution.

The Evolution of the Genetic Code

tRNA evolution led to a breakthrough that made life possible – the genetic code. The origins of this lie way back in the polymer world, which is thought to have included GCG repeats, ACCA, and glycine, the simplest amino acid. Glycine attached to ACCA, forming ACCA-Gly. Annealing between ACCA-Gly and GCG repeats led to polyglycine formation – aided by primordial hydration-dehydration cycles. A relic of this is seen in tRNA's anticodon stem-loop-stem sequence CCGGG_CU/???AA_CCCGG (/ indicates a U-turn and _ separates distinct features). This suggests the presence of self-

replicating RNA systems in the polymer world. The 7 nt core of this sequence CU/???AA – the anticodon loop – became instrumental to the development of the genetic code. Arguably, this sequence is pre-life's most central intellectual property.

Polyglycine is a cross-linking agent that stabilised protocells – the precursor to cell membranes. The polymer world led to the minihelix world, then the tRNA world, selected as a progressively improved way to generate polyglycine. Minihelix and tRNA worlds also saw the development of primitive catalysts and ribozymes – RNA enzymes. These were needed to generate, replicate, and ligate RNA strands – instrumental in the transition to the tRNA world. This suite of ribozymes included ligases, helicases, chaperones, replicases and endonucleases – all predating protein enzymes.

Without tRNA, no genetic code would be possible. It is unlikely that an alternative to tRNA would have supported a genetic code. Starting with glycine, tRNA learned to code an expanding repertoire of amino acids. tRNA-polyglycine world transitioned to tRNA-GADV, and then to the 8 amino acid tRNA-GADVLSER world (G=glycine; A=alanine; D=aspartate; V=valine; L=leucine; S=serine; E=glutamate; R=arginine). The evolution of mRNA codons conformed to the tRNA anticodons – leading to the first sequence-dependent proteins. These included ribosomal proteins, AARS enzymes, $(\beta-\alpha)_8$ barrels (TIM barrels), $(\beta-\alpha)_8$ sheets (Rossmann folds), RNA modification enzymes, RNA and DNA polymerases – all coevolving with the genetic code. Translation of Rossman folds and barrel proteins – with metabolic roles – is thought to have arisen from the ligation of multiple identical RNAs. Once the genetic code evolved, true life began!

Professors Burton and Lei solved tRNA evolution by inspection using available online tools. As a puzzle, anyone can verify their work. As tRNA evolution is a problem in biological coding, learning and problem-solving, it ought to be of interest to gamers, programmers, and psychologists.

Learning to Code is a Self-Taught Process

Like a self-learning computer algorithm, the genetic code retrained itself over time. To accommodate newly encoded amino acids, older amino acid-anticodons were abandoned, though they still retain the most favoured anticodon positions (the genetic code is degenerate – multiple codons encode the same amino acid). As early ribosomes learned to read the genetic code, tRNA-34 and -36 – anticodon positions – were thought to be subject to wobbling (non-canonical base pairing). However, wobbling at tRNA-36 was suppressed with modifications to tRNA-37 and ribosomal 16S rRNA.

The Clearest History of Chemical Evolution to Date

Professors Burton and Lei's solving of tRNA evolution is highly significant. With their highly explanative theorem, they have related the clearest history to date of the chemical evolution of life on Earth. tRNA's self-taught coding journey kickstarted life and had profound implications for planet Earth for aeons to come.

Meet the Researcher

Professor Emeritus Zachary Frome Burton
Department of Biochemistry and Molecular Biology
Michigan State University
East Lansing, MI
USA

Professor Emeritus Zachary Burton graduated with a BSc and a PhD from the University of California, Los Angeles. After obtaining his PhD in 1980, he undertook postdoctoral research at the University of Wisconsin's McArdle Laboratory for Cancer Research and the University of Toronto's Charles H Best Laboratory. In 1987, he joined Michigan State University (MSU), where he rapidly rose through the academic ranks. Professor Burton's past research interests have included transcription factors in cancer and gene expression in disease. Over time, his focus has shifted to the evolution of transcription and translation systems, with a particular interest in tRNA and the genetic code. While Professor Burton retired from MSU in 2017, he has continued to publish exciting new insights into the origin of life.

CONTACT

E: burton@msu.edu
W: www.canr.msu.edu/people/burton_zachary

KEY COLLABORATORS

Professor Lei Lei, University of New England